

ABSTRACT

An index term extraction device including: input means (1) for inputting a document-to-be-surveyed d and documents-to-be-compared P ; index term extraction means (120) for extracting an index term from the document-to-be-surveyed d ; first appearance frequency calculation means (142) for calculating a function value $IDF(P)$ of the appearance frequency of the extracted index term in the documents-to-be-compared P ; similar documents selecting means (160) for selecting similar documents S similar to the document-to-be-surveyed d in the documents-to-be-compared P according to the data on the document-to-be-surveyed d ; second appearance frequency calculation means (171) for calculating the function value $IDF(S)$ of the appearance frequency of the extracted index term in the similar documents S ; and output means (4) for outputting each index term and its positioning data according to the combination of the function values of the respective appearance frequencies in the documents-to-be-compared and the similar documents which have been calculated. Thus, it is possible to accurately grasp the feature of the document-to-be-surveyed.